Volume 13 Number 24 1985 Nucleic Acids Research

Molecular cloning of human apolipoprotein B cDNA

P.Carlsson, S.O.Olofsson, G.Bondjers, C.Darnfors, O.Wiklund and G.Bjursell

Departments of Medical Biochemistry and Medicine I, University of Göteborg, P.O. Box 33031, S-400 33 Göteborg, Sweden

Received 4 September 1985; Revised and Accepted 21 November 1985

ABSTRACT

In this paper we describe the isolation of cDNA clones which code for parts of apolipoprotein B (apoB). The clones were obtained by immunoscreening of an expression library (λgt 11) derived from a human hepatoma cell line (Hep G2). The relationship between positive clones and apoB was established with immunochemical techniques using polyclonal as well as monoclonal antibodies. Recombinants, expressing nonoverlapping regions of apoB are described, all hybridizing with a very large mRNA (approximately 20.000 bases long). The nucleotide sequence obtained predicts a primary protein structure with a composition suitable for the formation of stretches of an amphipatic α-helix.

INTRODUCTION

Apolipoprotein B (apoB) is one of the major protein components of the plasma lipoproteins. One form of apoB, referred to as apoB 100 (1) is synthesized in the liver and consitutes the major protein component of the Low- and Very Low Density Lipoproteins (LDL and VLDL). ApoB 100 is synthesized as one polypeptide (2, 3) and the reported values on its molecular mass vary between 250 and 500 kDal (1, 4 - 8).

The protein is of great biological importance since it not only stabilizes VLDL and LDL particles but also is essential for their formation (9). In addition apoB as the ligand for the LDL receptor is involved in VLDL and LDL catabolism (10, 11).

ApoB is also of clinical importance since elevated plasma levels of apoB and LDL are associated with an increased risk for the development of complications to atherosclerosis (12, 13, 17). Increased levels of apoB (and LDL) in plasma can be caused by a decreased catabolism, due to LDL-receptor deficiency, or it could be the result of an increased synthesis of apoB (15). While the structure and biology of the LDL-receptor is known in great detail (10, 11, 16), fairly little is known about the structure of apoB and the regulation of its synthesis. Since apoB is a very large molecule and virtually insoluble in water, we believe that the sequence of the protein most easily can be revealed from the nucleotide sequence of the cDNA of apoB mRNA. The availability of such cDNA would also facilitate studies on the mechanism for the regulation of apoB synthesis.

Copy DNA clones can also be used to study RFLP of the apoB gene. Polymorphism has already been demonstrated with monoclonal antibody techniques (17). However, linkage between any apoB allele and hyperlipoproteinaemia has not been demonstrated so far.

Recently , two papers have appeared (18, 19) which describe partial apoB cDNA clones of human and rat origin, respectively. The clones described in this paper do not overlap with those published and hence add new information (749 amino acids.) on the structure of the human apoB protein.

MATERIALS AND METHODS

mRNA fractionation and construction of the cDNA library

Polyadenylated, cytoplasmic RNA, isolated (21) from $6x10^8$ Hep G2 cells (21, 22) was sized by electrophoresis in a 1 % low gelling temperature agarose gel (Sea Plaque) containing methylmercuric hydroxide (23) and the fraction migrating slower than 28 S rRNA (5.500 bases) was recovered (24, 25). Approximately 0.5 ug of this mRNA, denatured with methylmercuric hydroxide was primed with random sequence hexadeoxynucleotides (Pharmacia, Uppsala, Sweden) and reversely transcribed, using standard procedures (26). The second strand was synthesized as described by Gubler and Hoffman (27), except that DNA ligase and β-NAD was omitted and the reaction was carried out at 15°C. Excess RNA was digested with RNase A (50 ug/ml). The double-stranded cDNA was treated with EcoRI methylase (28) and, to produce blunt ends, with T4 DNA polymerase (29). Following ligation to EcoRI linkers and cleavage with EcoRI the cDNA was fractionated on a Sepharose 4B-CL column. The void volume fraction was precipitated with ethanol and the cDNA inserted into the EcoRI-site of λ gt 11 (30). In vitro packageing yielded $>10^7$ plaque forming units, about 20 % of which were recombinants as judged from the lac Z-phenotype on X-gal plates.

Immunoscreening of the cDNA-library

To screen the cDNA library for clones expressing parts of apoB, $5x10^5$ plaques were plated at a density of $2{,}5x10^4$/90 mm dish and probed with antiserum as described by Young and Davis (31). The antiserum, a mixture of polyclonal antibodies to apoB 100 and LDL-2 respectively, was passed over a column of human serum albumin coupled to Sepharose 4B-Cl. The unretained fraction obtained was used after dilution 1:100 and preincubation with a lysate of the host strain, E. coli Y1090 (31). Antibody binding was detected with ^{125}I-protein A (2μCi/90 mm filter), and autoradiography for 16 hr.

Immunological characterization of apoB-clones

Dot-blot. After plaque purification, each clone was dotted on a lawn of host bacteria and then grown, induced (31) and probed with the polyclonal antiserum to apoB 100 adsorbed as described above. When monoclonal antibodies were tested for reaction

with the λapoB clones, the filters were allowed to react with 5 ml of hybridoma culture medium, followed by a secondary antibody (rabbit-anti-mouse) to mediate the binding of protein A. With these two exceptions the procedure used was essentially that described in ref. 31.

Western blot. Lysogenized E. coli Y1089, grown and induced as described by Young and Davis (40) and recovered by pelletation was suspended in 50 mM Tris-HCl pH 6.8 with 1.5 % SDS, 50 mM DTT, 4 mM Urea and 1 mM PMSF (at 1.5 % the original volume) and lysed in a Dounce homogenizer. After being boiled for 5 minutes, the lysates were centrifuged at 12.000xg for 30 minutes at +4°C, and 50 ul of the clear supernatant was electrophoresed in a polyacrylamide gradient (3-15 %) slab gel containing SDS (32). The gels were either stained with Coomassie Brilliant Blue R 250, or blotted onto nitrocellulose filters as described earlier (33). When polyclonal antiserum to apoB 100 was used, the blots were incubated for 8 hours with the antiserum (adsorbed with human serum albumin and lysate of the host strain as described above) in 10mM Tris-HCl pH 7.4, 150 mM NaCl, 3 % BSA. This was followed by (i) an 8 hour wash with 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 1 % Triton X100, (ii) an 8 hour incubation with 10 uCi ^{125}I-labelled protein A in 10 mM Tris-HCl pH 7.4, 150 mM NaCl, 3% BSA and (iii) an 8 hour wash with 10 mM Tris-HCl pH 7.4, 0.1 % SDS, 0.1 % Triton X-100.

Essentially the same procedure was used for monoclonal antibodies with the exception that hybridoma cell culture medium (pH 7.4) with 3 % BSA was used and a second antibody, rabbit-anti-mouse IG (in 10 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 % BSA) was used to mediate the binding to protein A. The rabbit-anti-mouse IG was adsorbed with a lysate of the host strain (see above).

Northern blot. Cytoplasmic, polyadenylated RNA from 3x10^6 Hep G2 cells was fractionated on a 0.7 % agarose/formaldehyde gel (34) and blotted onto nitrocellulose (35). The blots were probed with pBR322-subclones of the apoB recombinants, labelled with ^{32}P dCTP by nick translation. Posthybridization washes were performed with high stringency (0.1 % SDS, 0.1 x SSC, 65°C).

Nucleotide sequence

cDNA inserts from the λ apoB recombinants were cut out with EcoRI and cloned in M13mp8 in both orientations. The nucleotide sequence was determined by the dideoxy/chain-termination method (36). To obtain sequence from both strands of the largest clone, λ apoB 10, this insert was further subcloned after digestion with PstI, MspI or Sau3A.

Protein sequence prediction

To predict the corresponding amino acid sequence of the cDNA-inserts, codon usage of the different reading frames were compared according to Staden (37). The predicted

primary structure was searched for hydrophilic epitopes as described by Hopp & Woods (38) and the relative probabilities for secondary structures (α-helix, β-sheet and turn) were calculated as described by Chou & Fasman (39). Charge distribution of the suggested α-helices was examined by molecular model building.

RESULTS AND DISCUSSION

The strategy for cloning cDNA, corresponding to the human apoB gene, was based on the following facts: (i) apoB is synthesized in a human hepatoma derived cell line, Hep G2 (3), thereby being a source of cytoplasmic RNA containing the apoB messenger. (ii) The estimated molecular mass of apoB suggests that the corresponding mRNA consists of at least 7.500 bases. Hence, size fractionation of total mRNA should increase the frequency of apoB mRNA, since very few mRNA species are expected in such a size range. (iii) Only scanty information exists concerning the amino acid sequence of apoB. On the other hand, apoB has been extensively studied with immunochemical techniques. Therefore, the availability in our laboratory of well characterized antibodies made us choose an expression vector system for the detection of apoB-cDNA clones.

Isolation and immunological characterization of cDNA clones

The expression library constructed as described in Methods, was screened with a mixture of two antisera, raised against apoB 100 and LDL-2 respectively (Fig. 1 A and B). A total of 15 different clones, out of 19 detected (λapoB 1-19), were further characterized and found to react with monospecific, polyclonal antibodies to apoB 100 (Fig. 1 C). The reaction between the 15 clones and the antiserum was decreased if the antiserum had been adsorbed with apoB 100 coupled to Sepharose 4B (data not shown). Four of the clones (λapoB 3, 13, 15 and 18) also reacted with a monoclonal antibody to apoB 100, a-B^{CVII} (Fig. 1 D). Since, in this cloning procedure, the detection of apoB cDNA clones is only based on immunological reactions, we would like to describe some properties of the antibodies used. The polyclonal antiserum to apoB 100 (33) has recently been used to immunoprecipitate apoB synthesized in Hep G2 cells and the results obtained indicate that the antibodies only precipitate apoB 100 and apoB 100 nascent polypeptides with estimated molecular masses larger than 100 kDa (3). The monoclonal antibody, a-B^{CVII} isolated as described earlier (33), reacts with LDL-2 in an ELISA but fails to react with HDL, albumin and other apolipoproteins. Furthermore, Western blot analysis showed a specific reaction between the antibody and apoB 100.

In λ gt 11 the introduced sequences are translated, fused to β-galactosidase (114 kDa). After electrophoresis in SDS/polyacrylamide gels, the hybrid proteins are expected to migrate slower than the β-galactosidase. When lysates of induced apoB-lysogens were analyzed by this method, such hybrid proteins were seen (Fig. 2 A). The estimated molecular masses of those hybrid proteins were roughly 140 kD, which is in

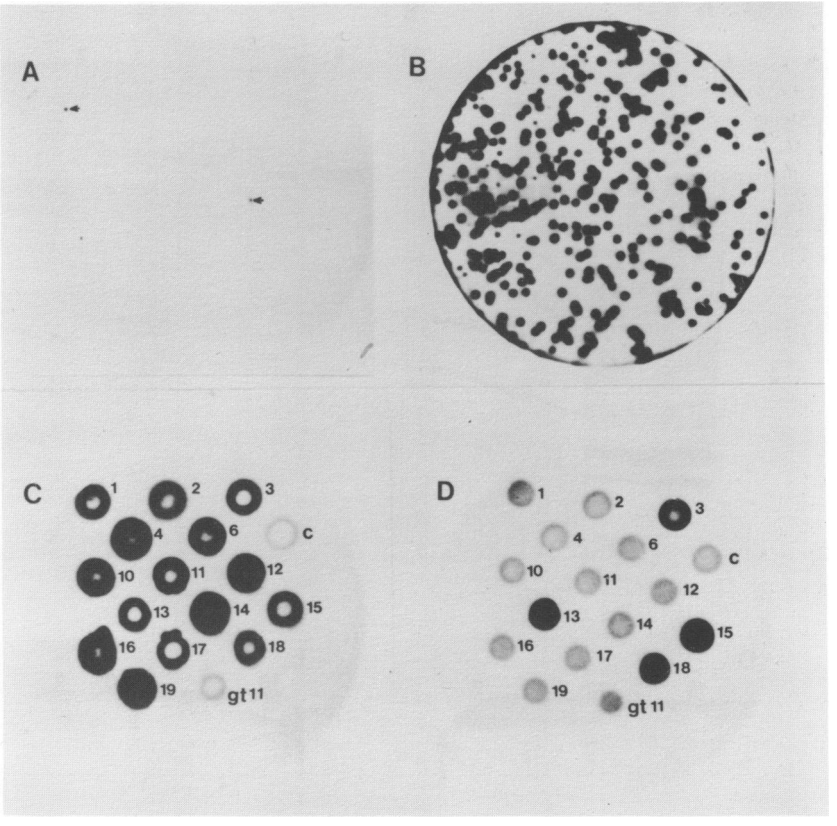


Figure 1. Screening of the cDNA library with antibodies. - (A) A 90 mm dish with $2{,}5 \times 10^4$ plaques. Arrows show two positive clones. (B) One of the clones in (A) after plaque purification. (C) Reaction of the 15 positive clones (λapoB 1 - λ apoB 19), the vector (λ gt 11) and a negative clone, picked from the library (c) with polyclonal antiserum to apoB 100. Serum from a nonimmunized rabbit did not react with any of the clones. (D) Reaction of the same clones and controls as in (C), with a monoclonal antibody to apoB 100, a-B^{cVII}. A monoclonal antibody synthesized by non-hybridized P3-X63-AG8 myeloma cells which was used in a parallel experiment did not react with any of the clones.

accordance with the expected values, calculated from the sizes of the cDNA-inserts. Western blot analysis confirmed that the binding of both the polyclonal and, in the case of λapoB 3, 13, 15 and 18, the monoclonal antibodies are confined to the hybrid proteins (Fig. 2 B-D).

Nucleotide sequence

As can be seen in Fig. 3 the cDNA clones were further characterized by nucleotide sequence analysis (36) and the restriction map predicted by the DNA sequence was

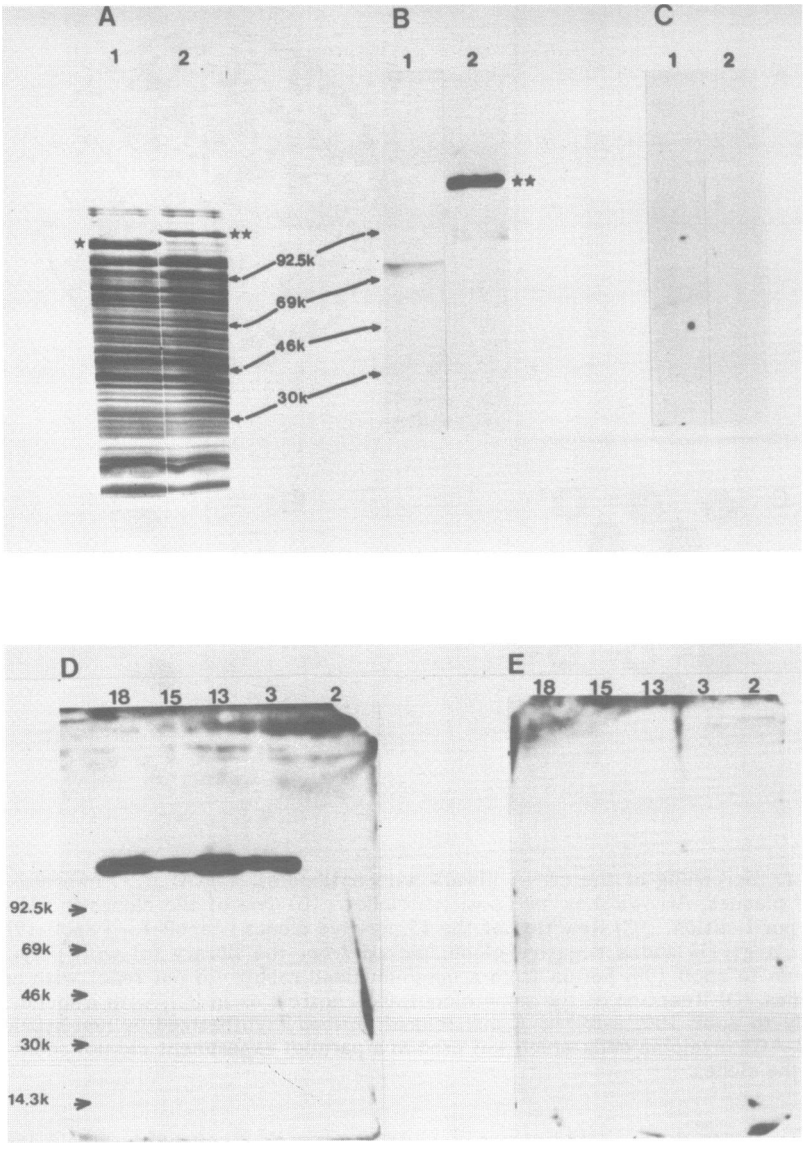
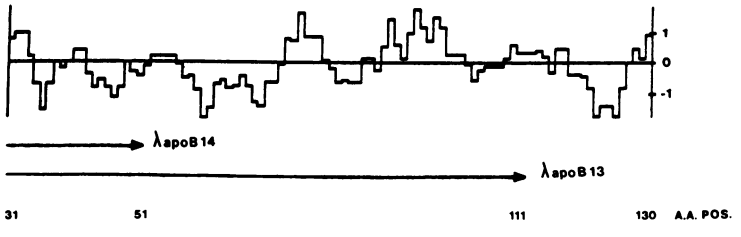


Figure 2. (A) Electrophoresis, in polyacrylamide gradient (3-15%) gels in the presence of SDS, of proteins present in lysate of E. Coli Y1089 (5) lysogenized with λ gt 11 (Lane 1) or one of the 15 λ apoB-clones (Lane 2). (*) indicates the position of β-galactosidase, while (**) indicates the position of the hybrid protein. (B) Western blot of the electrophoretograms shown in figure 2A with polyclonal antibodies to apoB 100. (C) The reaction between the filters (Fig.2B) and serum from nonimmunized rabbits. (D) Western blot of lysate from E. coli Y1089 lysogenized with clones λ apoB 2, 3, 13, 15 or 18 with a-B^{cVII}. (E) The reaction between the filter (Fig.2D) and monoclonal antibody synthesized by a non-hybridized myeloma cell line, p3x63-AG-68.

A

```
HisMetLysValLysHisLeuIleAspSerLeuIleAspPheLeuAsnPheProArgPhe
CATATGAAAGTCAAGCATCTGATTGACTCACTCATTGATTTTCTGAACTTCCCCAGATTC
        10        20        30        40        50        60

GlnPheProGlyLysProGlyIleTyrThrArgGluGluLeuCysThrMetPheIleArg
CAGTTTCCGGGGAAACCTGGGATATACACTAGGGAGGAACTTTGCACTATGTTCATAAGG
        70        80        90       100       110       120

GluValGlyThrValLeuSerGlnValTyrSerLysValHisAsnGlySerGluIleLeu
GAGGTAGGGACGGTACTGTCCCAGGTATATTCGAAAGTCCATAATGGTTCAGAAATACTG
       130       140       150       160       170       180

PheSerTyrPheGlnAspLeuValIleThrLeuProPheGluLeuArgLysHisLysLeu
TTTTCCTATTTCCAAGACCTAGTGATTACACTTCCTTTCGAGTTAAGGAAACATAAACTA
       190       200       210       220       230       240

IleAspValIleSerMetTyrArgGluLeuLeuLysAspLeuSerLysGluAlaGlnGlu
ATAGATGTAATCTCGATGTATAGGGAACTGTTGAAAGATTTATCAAAAGAAGCCCAAGAG
       250       260       270       280       290       300

ValPheLysAlaIleGlnSerLeuLysThrThrGluValLeuArgAsnLeuGlnAspLeu
GTATTTAAAGCCATTCAGTCTCTCAAGACCACAGAGGTGCTACGTAATCTTCAGGACCTT
       310       320       330       340       350       360

LeuGlnPheIlePheGlnLeuIleGluAspAsnIleLysGlnLeuLysGluMetLysPhe
TTACAATTCATTTTCCAACTAATAGAAGATAACATTAAACAGCTGAAAGAGATGAAATTT
       370       380       390       400       410       420

ThrTyrLeuIleAsnTyrIleGlnAspGluIleAsn
ACTTATCTTATTAATTATATCCAAGATGAGATCAAC
       430       440       450
```

B



C

```
GlnLeuProHisIleSerHisThrIleGluValProThrPheGlyLysLeuTyrSerIle
CAGCTTCCCCACATCTCACACACAATTGAAGTACCTACTTTTGGCAAGCTATACAGTATT
        10        20        30        40        50        60

LeuLysIleGlnSerProLeuPheThrLeuAspAlaAsnAlaAspIleGlyAsnGlyThr
CTGAAAATCCAATCTCCTCTTTTCACATTAGATGCAAATGCTGACATAGGGAATGGAACC
        70        80        90       100       110       120

ThrSerAlaAsnGluAlaGlyIleAlaAlaSerIleThrAlaLysGlyGluSerLysLeu
ACCTCAGCAAACGAAGCAGGTATCGCAGCTTCCATCACTGCCAAAGGAGAGTCCAAATTA
       130       140       150       160       170       180
```

```
GluValLeuAsnPheAspPheGlnAlaAsnAlaGlnLeuSerAsnProLysIleAsnPro
GAAGTTCTCAATTTTGATTTTCAAGCAAATGCACAACTCTCAAACCCTAAGATTAATCCG
        190         200         210         220         230         240

LeuAlaLeuLysGluTyrValLysPheSerSerLysTyrLeuArgThrGluHisGlySer
CTGGCTCTGAAGGAGTACGTGAAGTTCTCCAGCAAGTACCTGAGAACGGAGCATGGGAGT
        250         260         270         280         290         300

GluMetLeuPhePheGlyAsnAlaIleGluGlyLysSerAsnThrValAlaSerLeuHis
GAAATGCTGTTTTTTGGAAATGCCATTGAGGGAAAATCAAACACAGTGGCAAGTTTACAC
        310         320         330         340         350         360

ThrLysLysIleHisTrpArgLeuSerAsnGlyValIleValLysIleAsnAsnGlnLeu
ACAAAAAAAATACACTGGAGACTTAGTAATGGAGTGATTGTCAAGATAAACAATCAGCTT
        370         380         390         400         410         420

ThrLeuAsp
ACCCTGGATAG
        430
```

D

```
PheGluIleThrValProGluSerGlnLeuThrValSerGlnPheThrLeuProLysSer
TTTGAGATAACCGTGCCTGAATCTCAGTTAACTGTGTCCCAGTTCACGCTTCCAAAAAGT
        10          20          30          40          50          60

ValSerAspGlyIleAlaAlaLeuAspLeuAsnAlaValAlaAsnLysIleAlaAspPhe
GTTTCAGATGGCATTGCTGCTTTGGATCTAAATGCAGTAGCCAACAAGATCGCAGACTTT
        70          80          90          100         110         120

GluLeuProThrIleIleValProGluGlnThrIleGluIleProSerIleLysPheSer
GAGTTGCCCACCATCATCGTGCCTGAGCAGACCATTGAGATTCCCTCCATTAAGTTCTCT
        130         140         150         160         170         180

ValProAlaGlyIleValIleProSerPheGlnAlaLeuThrAlaArgPheGluValAsp
GTACCTGCTGGAATTGTCATTCCTTCCTTTCAAGCACTGACTGCACGCTTTGAGGTAGAC
        190         200         210         220         230         240

SerProValTyrAsnAlaThrTrpSerAlaSerLeuLysAsnLysAlaAspTyrValGlu
TCTCCCGTGTATAATGCCACTTGGAGTGCCAGTTTGAAAAACAAAGCAGATTATGTTGAA
        250         260         270         280         290         300

ThrValLeuAspSerThrCysSerSerThrValGlnPheLeuGluTyrGluLeuAsnVal
ACAGTCCTGGATTCCACATGCAGCTCAACCGTACAGTTCCTAGAATATGAACTAAATGTT
        310         320         330         340         350         360

LeuGlyThrHisLysIleGluAspGlyThrLeuAlaSerLysThrLysGlyThrPheAla
TTGGGAACACACAAAATCGAAGATGGTACGTTAGCCTCTAAGACTAAAGGAACATTTGCA
        370         380         390         400         410         420

HisArgAspPheSerAlaGluTyrGluGluAspGlyLysTyrGluGlyLeuGlnGluTrp
CACCGTGACTTCAGTGCAGAATATGAAGAAGATGGCAAATATGAAGGACTTCAGGAATGG
        430         440         450         460         470         480

GluGlyLysAlaHisLeuAsnIleLysSerProAlaPheThrAspLeuHisLeuArgTyr
GAAGGAAAAGCGCACCTCAATATCAAAAGCCCAGCGTTCACCGATCTCCATCTGCGCTAC
        490         500         510         520         530         540

GlnLysAspLysLysGlyIleSerThrSerAlaAlaSerProAlaValGlyThrValGly
CAGAAAGACAAGAAAGGCATCTCCACCTCAGCAGCCTCCCCAGCCGTAGGCACCGTGGGC
        550         560         570         580         590         600

MetAspMetAspGluAspAspAspPheSerLysTrpAsnPheTyrTyrSerProGlnSer
ATGGATATGGATGAAGATGACGACTTTTCTAAATGGAACTTCTACTACAGCCCTCAGTCC
        610         620         630         640         650         660
```

```
SerProAspLysLysLeuThrIlePheLysThrGluLeuArgValArgGluSerAspGlu
TCTCCAGATAAAAAACTCACCATATTCAAAACTGAGTTGAGGGTCCGGGAATCTGATGAG
      670       680       690       700       710       720

GluThrGlnIleLysValAsnTrpGluGluGluAlaAlaSerGlyLeuLeuThrSerLeu
GAAACTCAGATCAAAGTTAATTGGGAAGAAGAGGCAGCTTCTGGCTTGCTAACCTCTCTG
      730       740       750       760       770       780

LysAspAsnValProLysAlaThrGlyValLeuTyrAspTyrValAsnLysTyrHisTrp
AAAGACAACGTGCCCAAGGCCACAGGGGTCCTTTATGATTATGTCAACAAGTACCACTGG
      790       800       810       820       830       840

GluHisThrGlyLeuThrLeuArgGluValSerSerLysLeuArgArgAsnLeuGlnAsn
GAACACACAGGGCTCACCCTGAGAGAAGTGTCTTCAAAGCTGAGAAGAAATCTGCAGAAC
      850       860       870       880       890       900

AsnAlaGluTrpValTyrGlnGlyAlaIleArgGlnIleAspAspIleAspValArgPhe
AATGCTGAGTGGGTTTATCAAGGGGCCATTAGGCAAATTGATGATATCGACGTGAGGTTC
      910       920       930       940       950       960

GlnLysAlaAlaSerGlyThrThrGlyThrTyrGlnGluTrpLysAspLysAlaGlnAsn
CAGAAAGCAGCCAGTGGCACCACTGGGACCTACCAAGAGTGGAAGGACAAGGCCCAGAAT
      970       980       990      1000      1010      1020

LeuTyrGlnGluLeuLeuThrGlnGluGlyGlnAlaSerPheGlnGlyLeuLysAspAsn
CTGTACCAGGAACTGTTGACTCAGGAAGGCCAAGCCAGTTTCCAGGGACTCAAGGATAAC
     1030      1040      1050      1060      1070      1080

ValPheAspGlyLeuValArgValThrGlnGluPhe
GTGTTTGATGGCTTGGTACGAGTTACTCAAGAATTC
     1090      1100      1110
```

E

```
ProLeuAlaPheThrPheSerHisAspTyrLysGlySerThrSerHisHisLeuValSer
CCTCTGGCATTTACTTTCTCTCATGATTACAAAGGCTCCACAAGTCATCATCTCGTGTCT
        10        20        30        40        50        60

ArgLysSerIleSerAlaAlaLeuGluHisLysValSerAlaLeuLeuThrProAlaGlu
AGGAAAAGCATCAGTGCAGCTCTTGAACACAAAGTCAGTGCCCTGCTTACTCCAGCTGAG
        70        80        90       100       110       120

GlnThrGlyThrTrpLysLeuLysThrGlnPheAsnAsnAsnGluTyrSerGlnAspLeu
CAGACAGGCACCTGGAAACTCAAGACCCAATTTAACAACAATGAATACAGCCAGGACTTG
       130       140       150       160       170       180

AspAlaTyrAsnThrLysAspLysIleGlyValGluLeuThrGlyArgThrLeuAlaAsp
GATGCTTACAACACTAAAGATAAAATTGGCGTGGAGCTTACTGGACGAACTCTGGCTGAC
       190       200       210       220       230       240

LeuThr
CTAACT
```

Figure 3. Nucleotide sequence and predicted amino acid sequences of cDNA inserts from λapoB 3 (A), λapoB 17 (C)λapoB 1 together withλapoB 10 (D) andλapoB 2 (E). A hydrophilicity plot (38) of the 100 amino acids from nucleotide 64 to 364 in the λapoB 3 sequence is shown in B. Arrows indicate the relative positions of λapoB 13 (reacting with a-B^{cVII}) and λapoB 14 (not reacting with a-B^{cVII}).

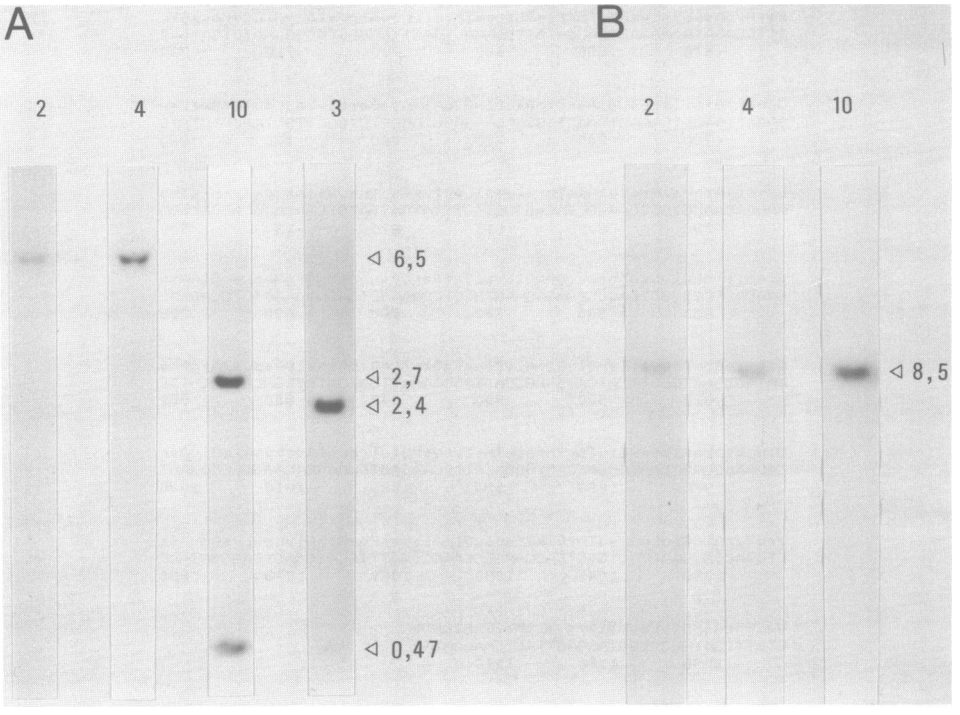


Figure 4. Southern blots of restriction enzymes digested total human DNA, probed with the different λ apoB clones. The DNA was digested with either (A) MspI or(B) PvuII. Numbers indicate the various λ apoB clones used as probes on the respective blots.

confirmed for PstI, EcoRV, PvuII, HincII, Sau3A and MspI. The sequences obtained revealed that the 15 clones could be divided into four different groups with overlapping sequences within each group. Group I, spanning a region of 456 bp, consists of λapoB 3, 13, 14, 15 and 18. Group II (spanning 431 bp) consists of λapoB 4, 6, 11, 12, 16, 17 and 19. Group III (spanning 1128 bp) consists of λ apoB 1 and 10 and finally group IV (spanning 246 bp) consists of λ apoB 2.

Organisation of the cDNA clones on the genome.

In order to analyse the organisation of the described λ apoB clones on the genome, total human DNA was digested with various restriction enzymes and probed with DNA representing the 4 different λapoB-groups. The data can be summarized from results obtained using restriction enzymes MspI and PvuII respectively. As can be seen in figure 4 λapoB 10 and 3 both detect a 0.47 kb long fragment. From the sequence it is clear that there is an MspI site 400 bp upstream in λapoB 10 and an MspI site 70 bp downstream in λapoB 3. Since λ apoB 10 ends with an EcoRI site and apoB 3 starts with

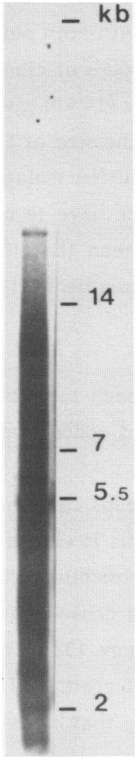


Figure 5. Northern blot of cytoplasmic, polyadenylated RNA from $3x10^6$ Hep G2 cells. The blot was probed with a pBR 322-subclone of λ apoB 3, Size markers are the rRNA's 18 S, 28 S and the rRNA precursors 32 S and 45 S. Extrapolation gives a size of approx. 20.000 bases of the apoB mRNA.

an EcoRI site these two clones must be situated next to each other with a common EcoRI site. Furthermore, it is clear from results obtained using PvuII that all clones detect a 8.5 kb long fragment. The sequence analysis reveals that there is a PvuII site in λapoB 2 and 3 respectively. Furthermore, λ apoB 2 and 4 detect a 6.5 kb long MspI fragment. By combinating the data the order of the clones at the genomic level can be obtained, and we conclude that the four λapoB clones are situated on a 8.5 kb long PvuII fragment in the order λ apoB 2, λ apoB 4, λapoB 10 ending with λ apoB 3, and λapoB 2 beeing closest to the 5' end of the mRNA.

Northern blot analysis

The three groups of cDNA-clones express peptides reacting with the same, monospecific antibodies, but shows no overlap in sequence. To ascertain that they are derived from the same mRNA-species, one member of each group (λapoB 2, 3, 4 and 10) were hybridized to blots of cytoplasmic, polyadenylated RNA from Hep G2-cells. They all showed hybridization to a very large messenger and to a smear, most likely representing degradation products of this mRNA. The results are exemplified by the

blot analysis obtained using λapoB 3 as a probe (Fig. 5). Combining this information with the observation that the clones contain hybrid proteins that react with both poly- and monoclonal antibodies to apoB, makes us conclude that the three groups of clones are all parts of the apoB gene. Extrapolation from the largest size marker (14 kb) gives an estimated size of approximately 20 kb for the apoB-mRNA. Hence the size of the human mRNA is in agreement with those recently published for apoB mRNA isolated from baboon and rat (18, 19). Mammalian messengers of such large size have to our knowledge not been described earlier. The size of the mRNA could mean that like some other large messengers it is built up by (40) or contains repetitive segments (16). No such data have been obtained so far (Fig. 3).

Predicted protein sequence

For all the three groups a single open reading frame was found. In each case this reading frame coincided with the frame selected by a computer-assisted codon-usage comparison (37).

The corresponding amino acid sequence are as shown in figure 3. All groups contained strongly hydrophilic domains, as revealed by hydrophilicity plots (38). This is exemplified in figure 4B for group I. This observation may be related to the possibility that hydrophilic regions are the most immunogenic portions of a protein and consequently will be preferentially selected by the immunological screening procedure (38). This possibility is supported by the following observation in the sequence of group I. The smallest clone of this group, λapoB 14 (which does not express the epitop that binds a-B^{cVII}) ends at position 153, while λ apoB 13 (the smallest clone that expresses the a-B^{cVII} epitope) extends 180 bases further, indicating that the epitope is localized between amino acid 51 and 111. This region contains the two most strongly hydrophilic domains within the sequence of group I (aa 74-79 GLU-LEU-ARG-LYS-HIS-LYS and aa 88-97 ARG-GLU-LEU-LEU-LYS-ASP-LEU-SER-LYS-GLU).

Structural studies of other apolipoproteions (for review with references see 41) have shown that the hydrophobic regions involved in the binding of lipid are present on regions of amphipatic α-helix. However, no such regions were observed in the rat apoB clones (19) and to investigate the possibility of amphipatic α-helices within apoB we carried out molecular model building of the region GLU_{58}-MET_{138} of group I (Fig. 4 A). This region has a high probability of α-helix structure, and indeed the results obtained (Fig. 6) indicate that this α-helix has pronounced amphipatic properties. At two positions (PRO_{72} and LYS_{109}) the amphipatic polarity turns 120° clockwise suggesting a turn in the helix at these positions. It is of course possible that this amphipatic α-helix constitutes a portion of a globular structure of apoB. Hence, more sequence data and results from lipid binding experiments are needed in order to evaluate the significance of amphipatic α-helix for the binding of apoB to lipids.

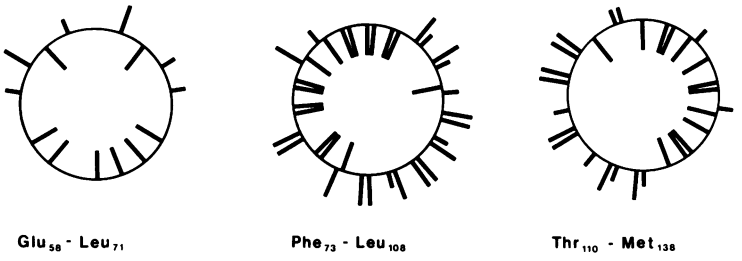


Figure 6. Schematic presentation of the amphipatic α-helix region within the λapoB 3-sequence. Each diagram is an end-on-view of a helical domain with bars representing amino acid side chains. Long bars pointing outwards designate charged side chains, short bars designate polar side chains and long bars pointing inwards designate hydrophobic side chains. The diagrams are drawn in phase, hence it is evident that at Pro_{72} and Lys_{109} the amphipatic polarity turn 120°clockwise, suggesting a bend in the helix at these positions.

To summarize, we conclude that we have cloned and characterized approximately 20 % of the coding regions of the unusually large mRNA for apolipoprotein B. In this context we would like to point out that the construction of our library, in contrast to those described (18, 19) involves fractionation of RNA and random priming of the reversed transcriptase reaction. This approach might turn out to facilitate the isolation of the 5' portion of such a long mRNA. Work is now in progress for a further characterization of the mRNA, the corresponding genomic DNA and polymorphism in this gene. Furthermore, we think that availability of the clones will be of a general interest in atherosclerosis research.

ACKNOWLEDGEMENTS

We are grateful to Maj-Lis Hermansson and Anita Magnusson for skilful technical assistance. This study was supported by grants from the Swedish National Science Research Council, the Swedish Medical Research Council (Projects Nos 7142 and 4531), the Swedish Oleo-Margarine Foundation of Nutritional Research, the Swedish National Association against Heart and Lung Diseases, the King Gustav V's and Queen Victoria's Foundation and AB Hässle (Göteborg, Sweden).

References

1. Kane, J.P., Hardman, D.A. and Paulus, H.E. (1980) Proc. Natl. Acad Sci. U.S.A. 77, 2465-2469.
2. Siuta-Mangano, P., Howard, S.C.. Lennarz, W.J. and Lane, D.M. (1982) J. Biol. Chem. 257, 4292-4300.
3. Wettesten, M., Boström, K., Bondjers, G., Jarfeldt, M., Norfeldt, P.-I., Carrella, M., Wiklund, O., Borén, J. and Olofsson, S.-O. (1985) Eur. J. Biochem. 149, 461-466.
4. Smith, R., Dawson, J.R. and Tanford, C. (1972) J. Biol. Chem. 247, 3376-3381.

5. Ikai, A. and Hasegawa, M. (1978) J. Biochem. 83, 755-759.
6. Steele, J.C.H., Jr and Reynolds, J.A. (1979) J. Biol. Chem. 254, 1639-1643.
7. Steele, J.C.H., Jr and Reynolds, J.A. (1979) J. Biol. Chem. 254, 1633-1638.
8. Elovson, J., Jacobs, J.C., Schumaker, V.N. and Puppione, D.L. (1985) Biochemistry 24, 1569-1578.
9. Malloy, M.J., Kane, J.P., Hardman, D.A., Hamilton, R.L. and Dalal, K.B. (1981) J. Clin. Invest. 67, 1441-1450.
10. Goldstein, J.L. and Brown, M.S. (1977) Ann. Rev. Biochem. 46, 897-930.
11. Brown, M.S., Anderson, R.G.W. and Goldstein, J.L. (1983) Cell 32, 663-667.
12. The lipid research clinical coronary primary prevention trial results II. The relationship of reduction in incidence of coronary heart disease to cholesterol lowering (1984) JAMA 251, 365-374.
13. Brunzell, J.D., Sniderman, A.D., Albers, J.J. and Kwiterovich Jr, P.O. (1984) Arteriosclerosis 4, 79-83.
14. Whayne, T.F., Alaupovic, P., Curry, M.D., Lee, E.T., Anderson, P.S. and Schechter, E. (1981) Atherosclerosis 39, 411-424.
15. Kesaniemi, Y.A. and Grundy, S.M. (1983) Arteriosclerosis 3, 40-46.
16. Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L. and Russell, D.W. (1984) Cell 39, 27-38.
17. Schumaker, V.N., Robinson, M.T., Curtiss, L.K., Butler, R. and Sparkers, R.S. (1984) J. Biol Chem. 259, 6423-6430.
18. Deeb S.S., Motulsky A.G. & Albers J.J. (1985) Proc. Natl. Acad. Sci. USA 82, 4983-4986.
19. Lusis A.J., West R., Mehrabin M., Reuben M.A., Leboeuf R.C., Kaptein J.S., Johnson D.F., Schumaker V.N., Yuhasz M.P., Schotz M.C. & Elovson J. (1985) Proc. Natl. Acad. Sci. USA 82, 4597-4601.
20. Favaloro, J., Treisman, R. and Kamen, R. (1980) Meth. Enzym. 65, 718-749.
21. Aden, D.P., Fogel, A., Plotkin, S.. Damjanow, I. and Knowles, B.B. (1979) Nature 282, 615-616.
22. Knowles, B.B., Howe, D. and Aden, D.P. (1980) Science 209, 497-499.
23. Bailey, J.M. and Davidson, N. (1976) Anal. Biochem. 70, 75-85.
24. Wieslander, L. (1979) Anal. Biochem. 98, 305-309.
25. Lemischka, I.R., Farmek, S., Racaniello, V.R. and Sharp, P.A. (1981) J. Mol. Biol. 151, 101-120.
26. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) In: Molecular Cloning, A laboratory manual (Cold Spring Harbor Laboratory, New York, 1982).
27. Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263-269.
28. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Conell, C. and Quon, D. (1978) Cell 15, 687-701.
29. O'Farrel, P. (1981) Bethesda Research Lab. Focus 3:3, 1-2.
30. Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad Sci. U.S.A. 80, 1194-1198.
31. Young, R.A. and Davis, R.W. (1983) Science 222, 778-782.
32. Laemmli, U.K. (1970) Nature 227, 680-685.
33. Boström, K., Wettesten, M., Wiklund, O., Bondjers, G., Lundholm, K., Elias, P., Norfeldt, P.-I. and Olofsson, S.-O. (1984) Eur. J. Biochem. 143, 101-107.
34. Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, A. (1977) Biochemistry 21, 4743-4751.
35. Thomas, P. (1983) Meth. Enzym. 100, 255-266.
36. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad Sci. U.S.A. 74, 5463-5467.
37. Staden, R. (1984) Nucleic Acids Res. 12:1, 551-567.
38. Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad Sci., U.S.A. 78, 3824-3828.
39. Chou, P.Y. and Fasman, G.D. (1978) Ann. Rev. Biochem. 47, 251-276.
40. Sümegi, I., Wieslander, L. and Daneholt, B. (1982) Cell 30, 579-587.
41. Smith, L.C., Pownall, H.J. and Gotto Jr, A.M. (1978) Ann. Rev. Biochem. 47, 751-777.